	COL 774	
Ayush Verma	Assignment -1 Report	2016CS10374

1 -

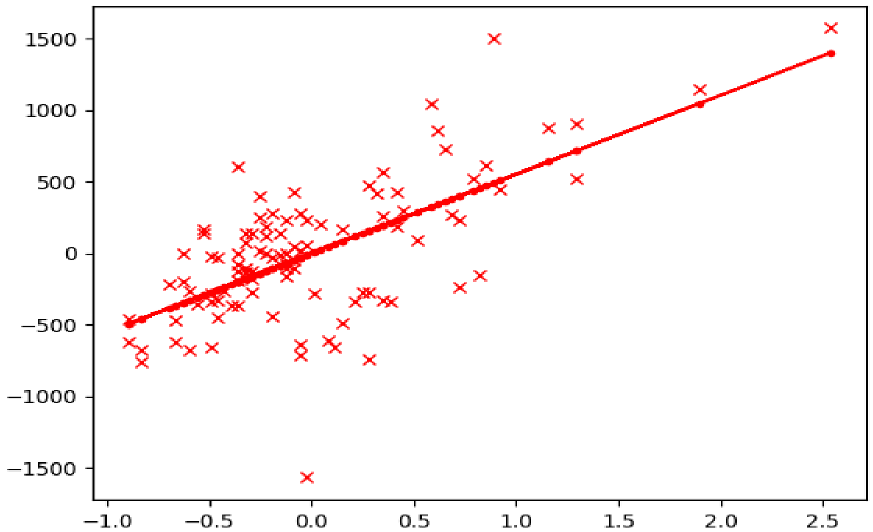
a) Batch Gradient descent

learning rate = 0.01

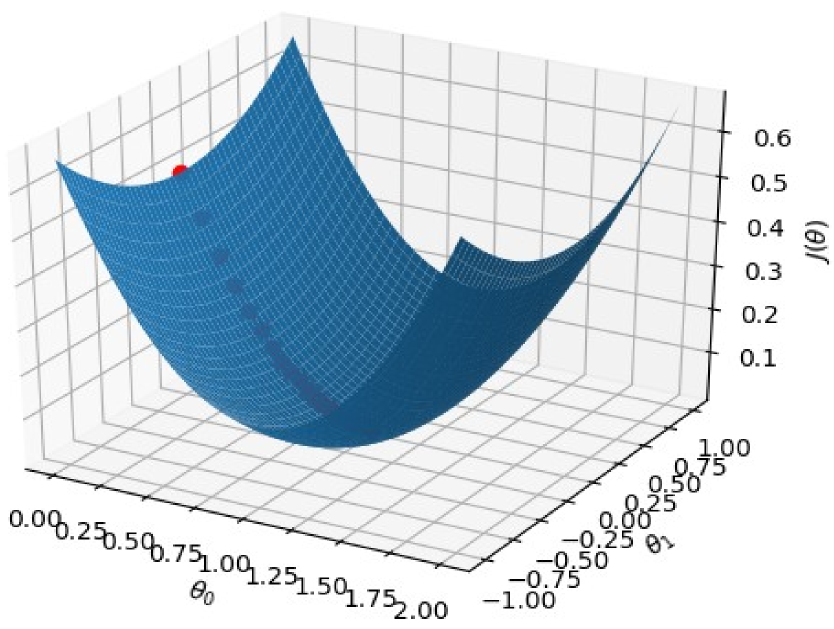
Stopping criteria = |J(Θ)_new - (Θ)_old | < 1e-7

Final Theta = [[0.9957635]
[0.00207673]]

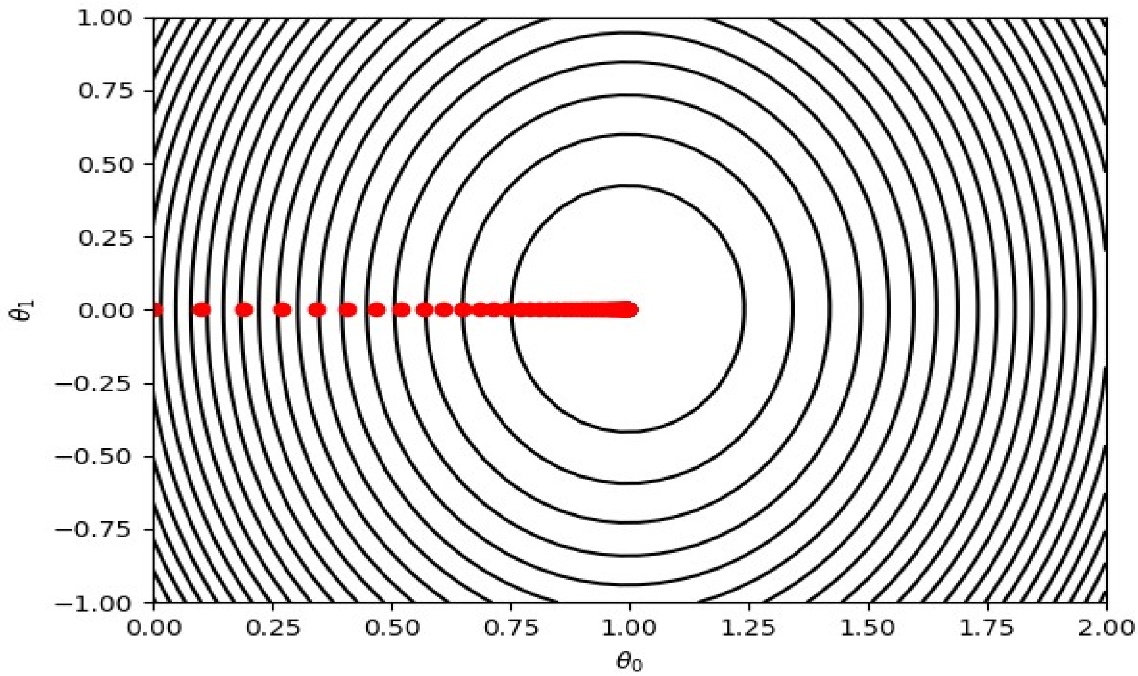
b)



c)

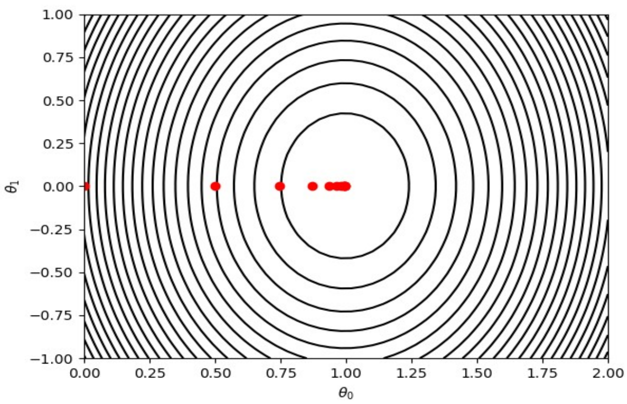


d) contour for eta =0.1

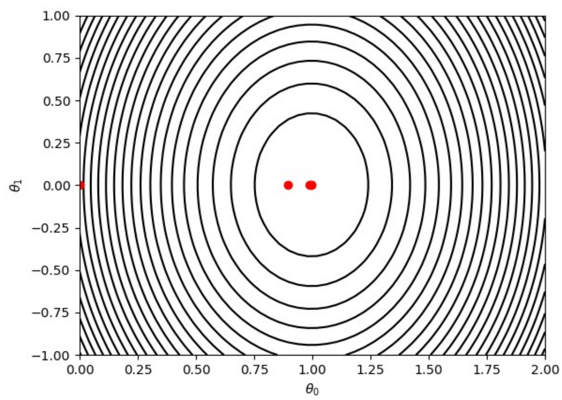


e)

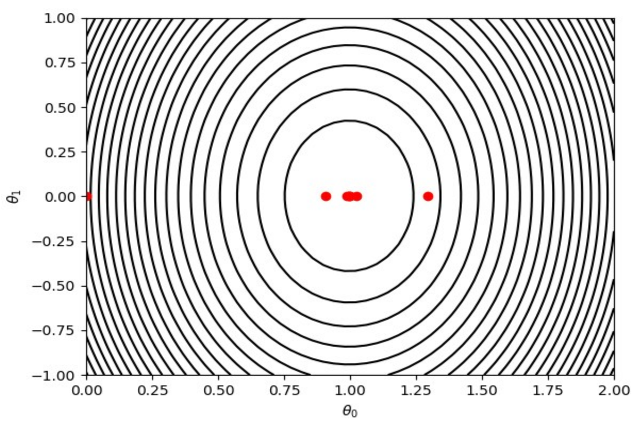
contour for eta =0.1



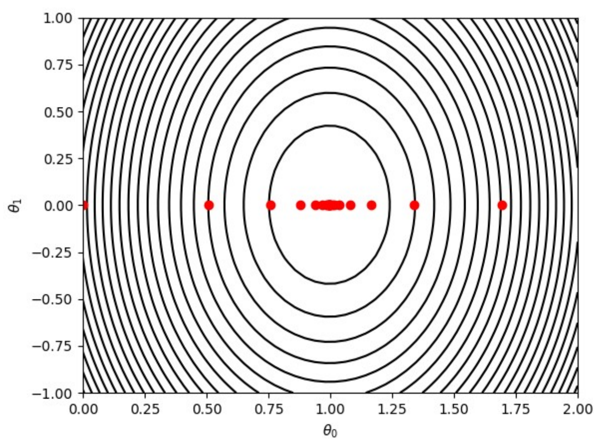
contour for eta = 0.9



contour for eta =1.3



contour for eta =1.7

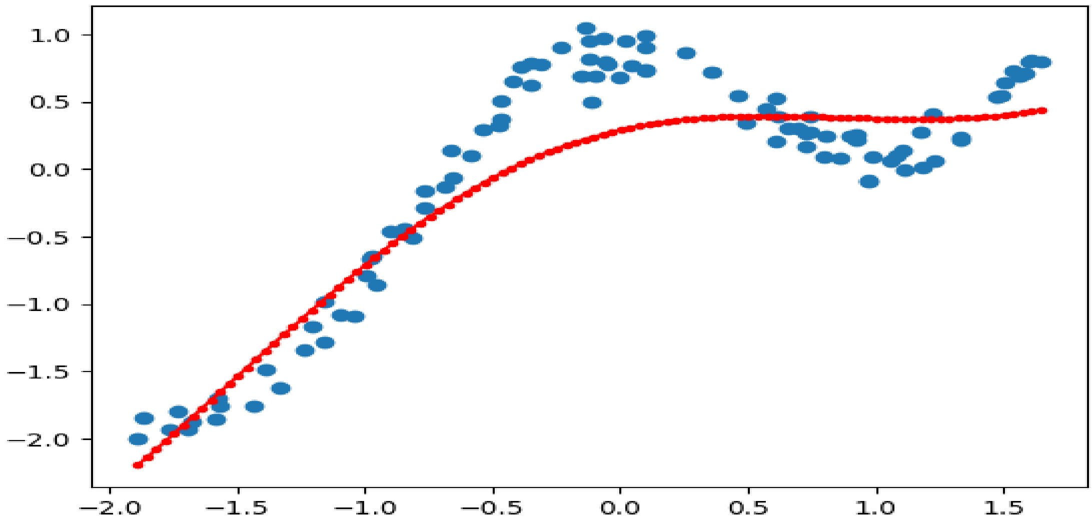


It is clear from above curves that beyond a certain point cost function starts diverging.So, smaller values of learning rates are preferred but not too small so that our hypothesis converges in finite number of steps.

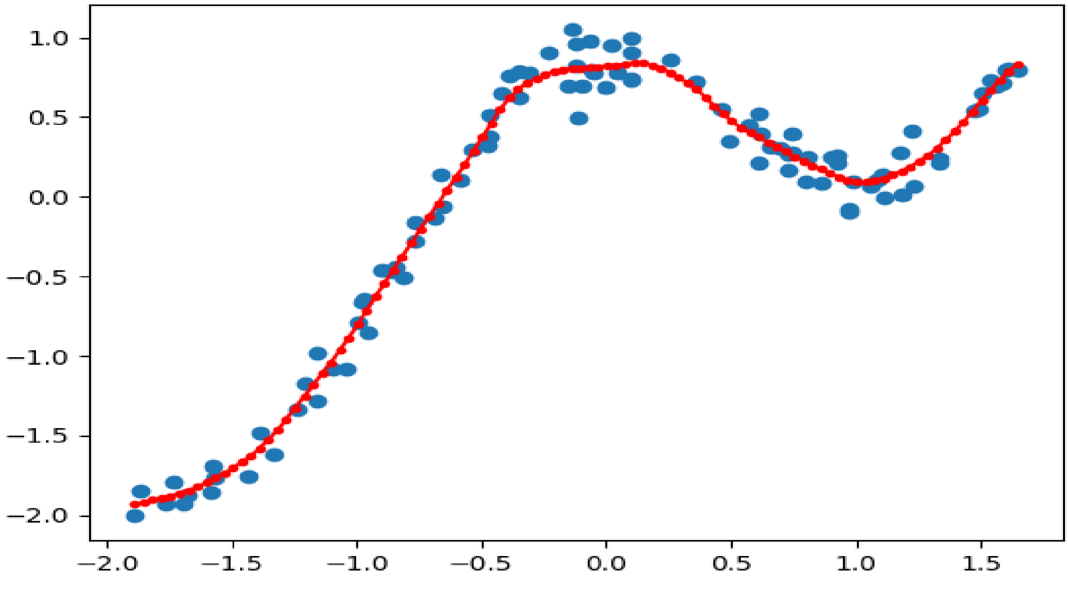
2 -
a) Unweighted Linear regression
    Theta = [1.021356716 , 3.86373461]
Linear regression model cannot fit the data as data itself is non-linear

b)

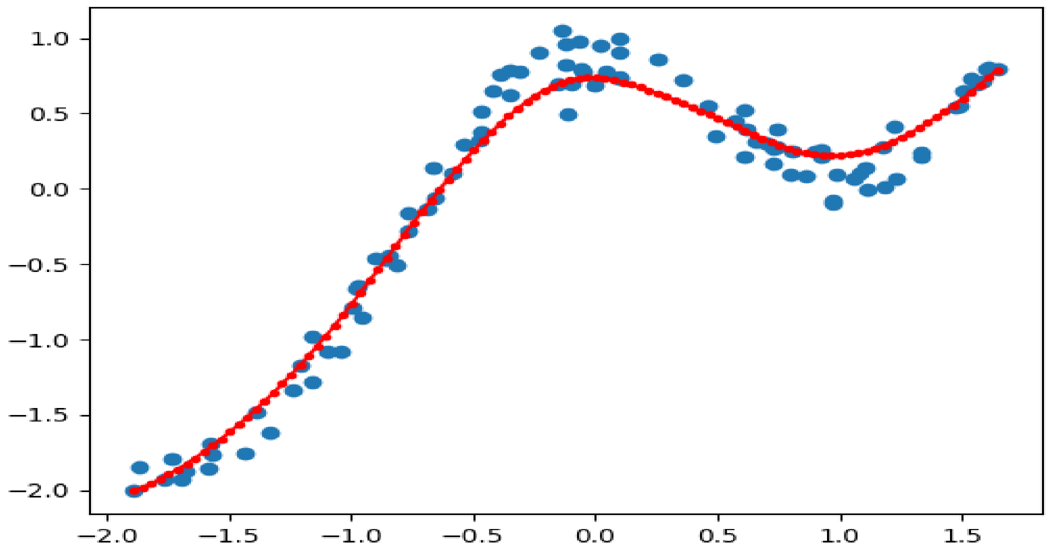
tau =0.8



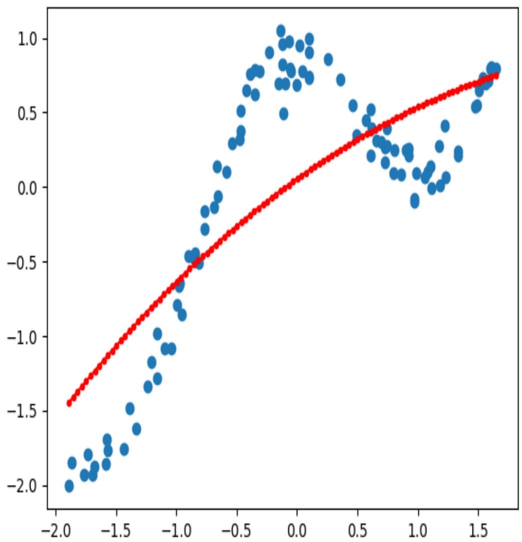
tau =0.1



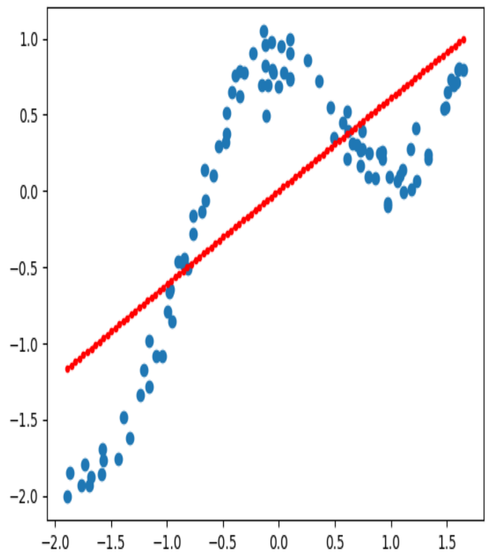
tau = 0.3



tau = 2



tau =10



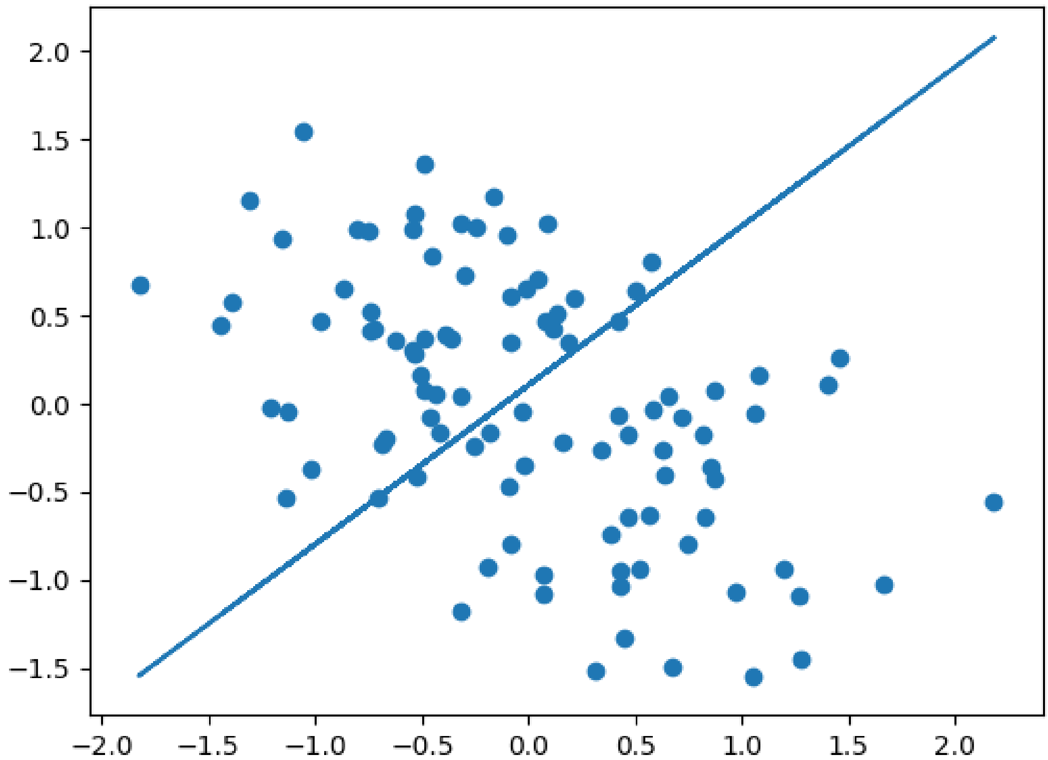
From above curve for various values of tau we can see that model that best fit with data is one with tau =0.8. If we decrease value of tau model overfits the data while increasing tau model become similar to linear regression. So for every new data we need to find optimal value of tau for every new set of data.

3 - Logistic Regression using Newton's Method

a) Theta = [[0.40125316]
[3.4141069]
[-3.78084345]]

Iteration = 8

b)



4 -

a) phi = 0.5

u0 = [[-0.75529433 0.68509431]]

u1 = [[0.75529433 -0.68509431]]

sigma = [[0.42953048 -0.02247228]
[-0.02247228 0.53064579]]

d) phi = 0.5

u0 = [[-0.75529433 0.68509431]]

u1 = [[0.75529433 -0.68509431]]

sigma0 = [[0.38158978 -0.15486516]
[-0.15486516 0.64773717]]

sigma1 = [[0.47747117 0.1099206]
[0.1099206 0.41355441]]

b)+c)+e)

Linaer decision boundary ($\Sigma 0 = \Sigma 1 = \Sigma$):

The decision boundary is given by the equation :

$AX + B = 0$

$A = 2 * (\mu_0^T \Sigma^{-1} - \mu_1^T \Sigma^{-1})$

$B = 2 * \log((1/\varnothing) - 1) + \mu_1^T \Sigma^{-1} \mu_1 - \mu_0^T \Sigma^{-1} \mu_0$

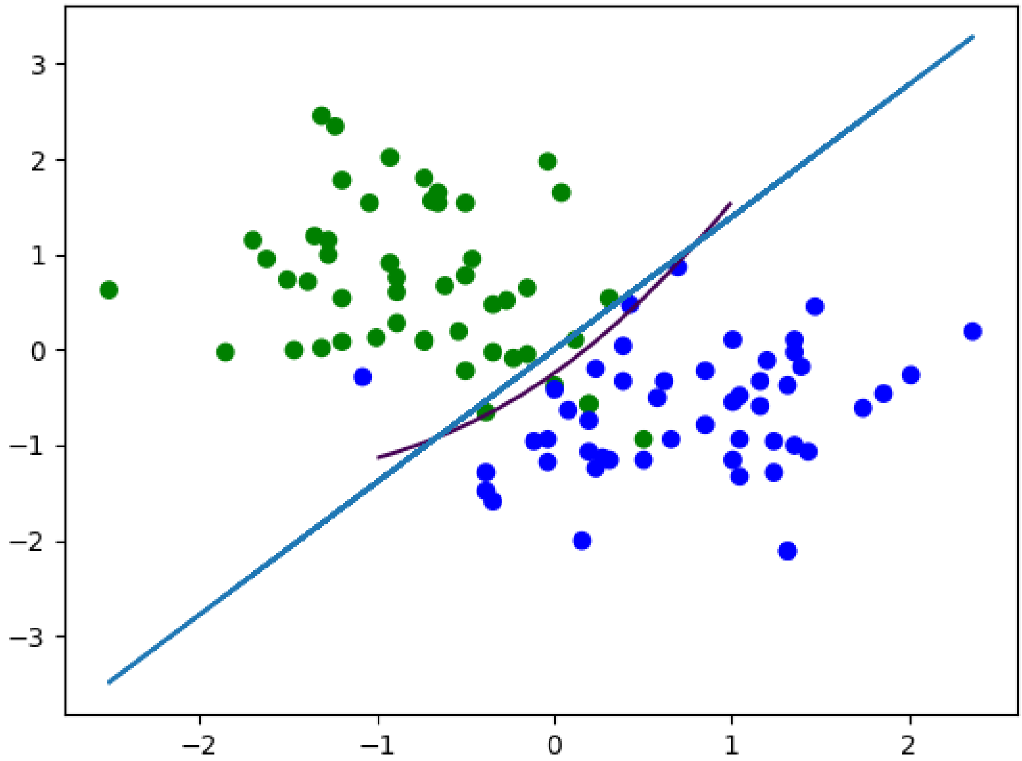
Quadractic Decision Boundary Equation

$X^T AX + BX + C = 0$

$A = \Sigma_0^{-1} - \Sigma_1^{-1}$

$B = -2 * (\mu_0^T \Sigma_0^{-1} - \mu_1^T \Sigma_1^{-1})$

$C = (\mu_0^T \Sigma^{-1}_0 \mu_0 - \mu_1^T \Sigma^{-1}_1 \mu_1 - 2 * \log((1/\varnothing) - 1)) * (|\Sigma_0 / \Sigma_1|)$



From the above plot we can see that the quadratic decision boundary is better method of classification than the linear decision boundary (equivalent to logistic regression). This means that the assumption that the underlying data belongs to Gaussian distribution is valid to a great extent.